

CLAIMS

What is claimed is:

1. A method for extracting at least one funnel from at least one input clickstream (CS), said CS representing an ordered path of web pages P_1 to P_N successively viewed by a user, said method comprising:

storing each ordered path within the CS in one of up to N tree structures, each of said tree structures having a root node P_i corresponding to one of pages P_1 to P_N and having successive child nodes corresponding to the successive pages after P_i of P_{i+1} to P_{i+X} , said X representing a depth criterion; and

extracting a list of stored paths from each tree structure from the root node P_i to each end node to represent a set of funnels corresponding to the CS.

2. The method of claim 1, wherein storing comprises incrementing a counter associated with each node in each ordered path as the ordered path is stored, and further comprising analyzing each stored path in each tree structure using the counters to identify the stored paths that satisfy one or more selected input criteria.

3. The method of claim 2, wherein the input criteria is selected from a group consisting of a width criterion, a starting page criterion, and an end page criterion, said width criterion representing a retention rate, said starting page criterion specifying a set of pages in CS from which P_i is selected, and said end page criterion specifying a set of pages in CS that can serve as end nodes in each tree structure.

4. The method of claim 3, wherein the retention rate is measured by a frequency relative to either the root node or the previous child node.
5. The method of claim 3, wherein the retention rate is measured by an absolute frequency.
6. The method of claim 3, wherein the width criterion specifies a minimum width or a maximum width or both.
7. The method of claim 1, wherein the depth criterion specifies a maximum depth.
8. The method of claim 1, wherein the depth criterion represents a desired number of pages in each stored path.
9. The method of claim 1, wherein P_1 corresponds to one of P_1 to P_N specified as starting page criterion.
10. The method of claim 1, wherein the stored path is a subpath.
11. The method of claim 1, wherein storing comprises scanning the CS one time.

12. The method of claim 1, further comprising automatically generating a report of the extracted list of stored paths.

13. One or more computer readable media having computer-executable instructions for performing the method recited in claim 1.

14. A method for analysis of funnels in at least one input clickstream (CS), said method comprising:

storing one or more paths within the CS satisfying a first input criterion in one or more tree structures; and

analyzing each tree structure to identify any of the paths that satisfy a second input criterion, said second input criterion representing a width criterion.

15. The method of claim 14, wherein the first input criterion comprises a depth criterion specifying a maximum number of pages in each stored path or a starting page criterion specifying a set of pages in the CS that can serve as a root node in one of the tree structures or both.

16. The method of claim 14, wherein the stored path is a subpath.

17. The method of claim 14, wherein storing comprises:

creating a tree structure for each starting page of the one or more paths; and

creating one or more branches in each tree structure for each path that starts with one of the starting pages.

18. One or more computer readable media having computer-executable instructions for performing the method recited in claim 14.

19. A method for extracting at least one funnel from one or more input clickstreams, each of said clickstreams representing an ordered path of web pages successively viewed by a user, said method comprising:

creating a tree structure for storing the ordered path for each of the clickstreams, each tree structure having a root node corresponding to the first page of the clickstream and having a child node corresponding to each of the successive pages in the clickstream;

for each input starting page, searching each tree structure to identify any stored path that starts at a node associated with the input starting page and satisfies an input depth requirement, said input depth requirement representing a desired number of pages in each path;

storing each identified path in a temporary tree; and

recurring through the temporary tree to identify any path that satisfies an input width requirement, said input width requirement representing a retention rate.

20. The method of claim 19, wherein the identified path is a subpath.

21. The method of claim 19, wherein the recursed path is a subpath.

22. The method of claim 19, wherein storing comprises incrementing a counter associated with each node in each path as the path is stored.

23. One or more computer readable media having computer-executable instructions for performing the method recited in claim 19.

24. A computer-readable medium having computer-executable components for extracting at least one funnel from at least one input clickstream (CS) representing an ordered path of successively viewed pages P_1 to P_N of a user, said components comprising:

a repository component for storing each ordered path within the CS in one of up to N tree structures, each of said tree structures having a root node P_i corresponding to one of pages P_1 to P_N and having successive child nodes corresponding to the successive pages after P_1 of P_{i+1} to P_{i+X} , said X representing a depth criterion;

a support component for incrementing a counter associated with each node in each ordered path as the ordered path is stored;

a funnel component for extracting a list of stored paths from each tree structure from the root node P_i to each end node to represent a set of funnels corresponding to the CS; and

a criteria component for analyzing each stored path in each tree structure using the counters to identify the stored paths that satisfy one or more input criteria.

25. The method of claim 24, wherein the input criteria is selected from a group consisting of a width criterion, a starting page criterion, and an end page criterion, said width criterion represents a retention rate, said starting page criterion specifies a set of pages in CS from which P_1 is selected, and said end page criterion specifies a set of pages in CS that can serve as end nodes in each tree structure.

26. A computer-readable medium having stored thereon a data structure for a particular node in a tree structure storing at least one click path from one or more input clickstreams representing an ordered path of successively viewed web pages P_1 to P_N of a user, said tree structure having a root node, each particular node associated with one of the viewed web pages, said data structure comprising:

a first field storing a page name representing a name of the viewed web page associated with the particular node; and

a second field storing a support value representing a frequency of appearance for a particular path starting with the root node of the tree structure and including the particular node.

27. The computer-readable medium of claim 26, wherein the data structure is created for each distinct web page in the input clickstreams as the clickstream is scanned.

28. A method for extracting at least one funnel from at least one input clickstream (CS) representing an ordered path of successively viewed pages P_1 to P_N , said method comprising:

reading through CS from P_1 to P_X , wherein X is less than or equal to N and represents an input depth;

creating a first tree with a root node associated with page P_1 , and with successive child nodes associated with pages P_2 to P_X , wherein P_X represents a child node with parent P_{X-1} ;

incrementing a counter associated with each node in the first tree as the node is created;

creating a second tree with a root node associated with page P_2 , and with successive child nodes associated with pages P_3 to P_{X+1} ;

incrementing a counter associated with each node in the second tree as the node is created;

creating additional trees rooted at each page P_{N-X+2} to P_N for all subpaths in CS starting with pages P_{N-X+2} to P_N ;

storing the subpaths that start at each page and ending at P_N in the respective tree so that new trees are created only when the trees or nodes have not already been created; and

running through all paths in each tree to extract and output only paths that satisfy input depth and input width criteria.

29. The method of claim 28, further comprising processing another clickstream according to the method described in claim 28.

30. The method of claim 28, wherein the tree for a page P_i is only created if P_i is part of an input starting page criterion.

31. One or more computer readable media having computer-executable instructions for performing the method recited in claim 28.